

Generation with Norsyg

Petter Haugereid

Department of Language, Literature, Mathematics and Interpreting
Western Norway University of Applied Sciences

Cambridge, July 16, 2019

Norsyg – Norwegian Syntax-based Grammar

Two central features of Norsyg

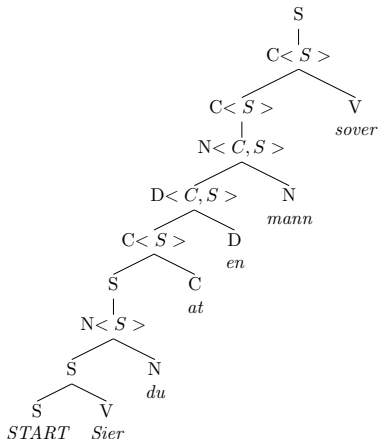
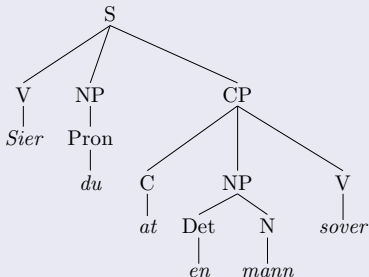
- 1 The grammar is strictly left-branching
 - ⇒ all rules are of the form Phrase ⇒ START/Phrase, (Word)
 - ⇒ compatible with incremental parsing
 - makes use of a *STACK* feature to account for constituent structure
- 2 The lexical entries do not introduce a relation on their *RELS* list
 - *PRED* values are specified in *KEYREL*
 - For verbs, function words, and parts of MWEs, the *PRED* value often is an underspecified type

Parse tree vs. constituent tree

Example

- (1) Sier du at en mann sover?
 Say you that a man sleeps
Are you saying that a man is asleep?

Constituent structure

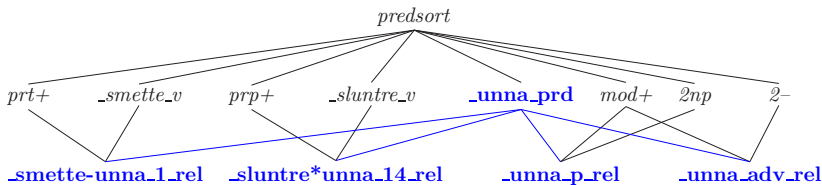


Unique lexical entries

Unique lexical entries are achieved by assuming subconstruction types:

- 1 Types are associated with category-specific rules
 - subject NPs
 - object NPs
 - objects introduced by complementizers
 - objects introduced by infinitival markers
 - selected prepositions
 - particles
 - adverbs/modifying prepositions
 - idiomatic nouns
- 2 Types are assigned to lexical items:
 - verbs
 - function words
 - words in idiomatic expressions
- 3 A hierarchy of atomic construction types determines what constructions can be formed

Simplified type hierarchy under *_unna_prd* ('away')



Lexical entry for the function word *unna* 'away'

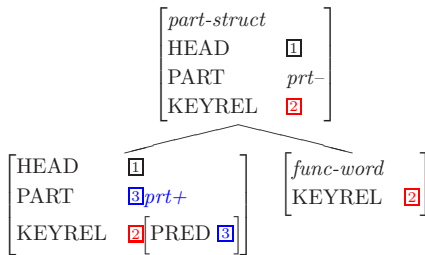
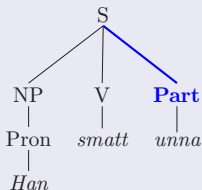
	<i>func-word</i>
ORTH	⟨"unna"⟩
CAT	[HEAD <i>adv-prep</i>]
CONT	[RELS ⟨!!⟩]
KEYREL	[PRED <i>_unna_prd</i>]

Tree with particle

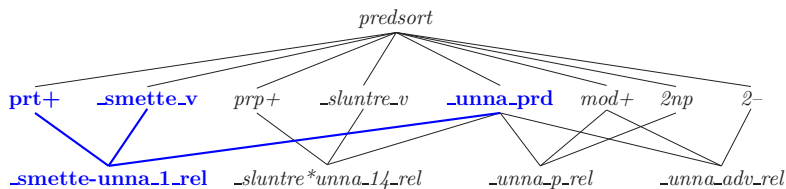
Example

(2) Han smatt **unna**.
 he dodged away(PRT)
He dodged.

Constituent structure



“unna” as particle

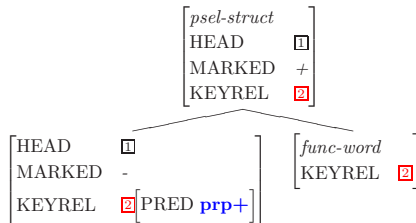
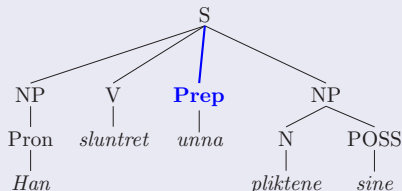


Tree with selected preposition

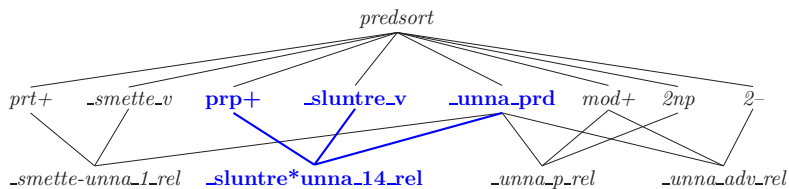
Example

(3) Han sluntret **unna**
 he idled away(Psel)
 pliktene sine.
 duties his
He shirked his duties.

Constituent structure



“unna” as selected preposition



Underspecified lexical entries

- 6000 lexical entries with underspecified PRED value
- 21000 argument frames

Challenges with generation

There are certain challenges with the design when it comes to generation

- 1 Since lexical entries have an empty RELS list, they are not indexed
- 2 The position where the predicate(s) of a constituent are added, matter
- 3 The position where subconstruction types are unified, matter

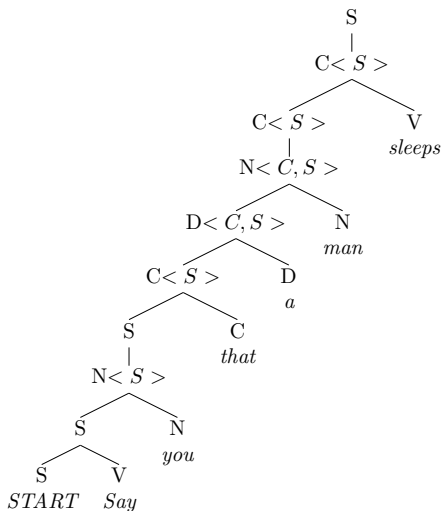
1. Lexical entry for the function word *unna* 'away'

<i>func-word</i>	
ORTH	⟨"unna"⟩
CAT	[HEAD <i>adv-prep</i>]
CONT	[RELS ⟨!!⟩]
KEYREL	[PRED <i>_unna_prd</i>]

Solution: A dummy LEXREL FEATURE

<i>func-word</i>	
ORTH	⟨"unna"⟩
CAT	[HEAD <i>adv-prep</i>]
CONT	[RELS ⟨!!⟩]
KEYREL	1 [PRED <i>_unna_prd</i>]
LEXREL	⟨! 1 !⟩

Parse tree – positions of entering PRED values and unifying subconstruction types



Results

- 266 of 272 items in the test suite generate
- Future work:
 - Punctuation is not handled well
 - Currently no handling of unknown words in generation